Estymacja przedziałowa - przedziały ufności

Próbę n-elementową charakteryzujemy jej parametrami (np. $\overline{x}$, s^2, s). Służą one do oceny wartości nieznanych parametrów populacji (m, σ^2, σ). Nazywamy je **estymatorami punktowymi** nieznanych parametrów populacyjnych.

Nieznany parametr populacji θ (np. EX=m, $D^2X=\sigma^2$) może być przybliżony przez swój estymator punktowy (np. $\overline{x}$, s^2, s). Może być też szacowany przy pomocy **przedziału ufności**. Estymacja przedziałowa polega na konstrukcji przedziału liczbowego, który z określonym z góry (bliskim 1) prawdopodobieństwem (poziomem ufności) będzie zawierał **nieznaną** wartość szacowanego **parametru populacji**.

Twórcą metody estymacji przedziałowej był statystyk polskiego pochodzenia Jerzy Spława-Neyman (1894-1981).

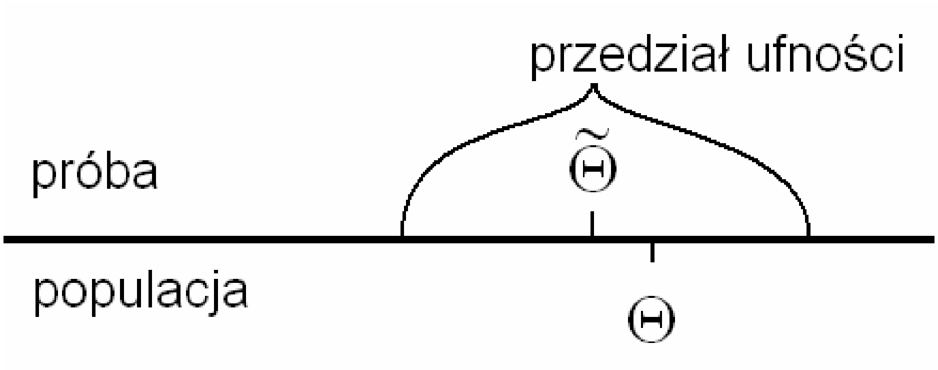
Postać przedziału ufności jest następująca:

$$\{g_1 \leq \Theta \leq g_2\}, \qquad P = 1 - a$$

Stwierdzamy, że z prawdopodobieństwem 1-α przedział ufności (g_1,g_2) zawiera szacowany parametr populacyjny Θ.

W tym zapisie Θ jest wielkością stałą, choć nieznaną, g_1 i g_2 zaś są wartościami liczbowymi wyznaczonymi z próby. Są one zmiennymi losowymi elementów próby, **statystykami z próby**.

Wielkość α to **poziom istotności** (lub **ryzyko błędu**, że określony na podstawie próby przedział nie zawiera parametru Θ), zaś 1-α to **poziom ufności**.



Dla rozkładu normalnego (nieznane s^2) przedział ufności dla średniej populacji ***m*** może mieć postać:

$$m \in (\bar{x} - t_{a,n-1}\frac{s}{\sqrt{n}};\ \bar{x} + t_{a,n-1}\frac{s}{\sqrt{n}}), \qquad P = 1 - a$$

Jest to najkrótszy przedział zawierający średnią populacyjną z założonym poziomem ufności. Wartość $t_{\alpha,n-1}$ odczytujemy z tablic rozkładu **t-Studenta**. Wartość α jest nazywana poziomem istotności, v=n–1 to liczba stopni swobody.

W pewnych przypadkach zależy nam na jednostronnych przedziałach ufności:

$$m \in (\bar{x} - t_{2a,n-1}\frac{s}{\sqrt{n}};\ \infty)$$

$$m \in (-\infty;\ \bar{x} + t_{2a,n-1}\frac{s}{\sqrt{n}})$$

Długość podanego przedziału dwustronnego dla średniej ***m*** wyraża formuła:

$$d = 2t_{a,n-1}\frac{s}{\sqrt{n}}$$

Zależy nam na tym, by przedział ufności dla wartości średniej ***m*** był jak najkrótszy. Możemy osiągnąć to:

- poprzez zwiększenie liczebności próby,
- zwiększając parametr α, a więc zmniejszając poziom ufności 1–α.

W pierwszym przypadku wiąże się to ze zwiększonym nakładem pracy i kosztów, w drugim – ze **zwiększeniem** ryzyka pomyłki (szacowany parametr nie będzie należał do określonego przez nas przedziału)

Jeśli chcemy oszacować parametr z określoną dokładnością *d*, to po odpowiednich przekształceniach wzorów na przedziały ufności możemy wyznaczyć liczebność próby losowej potrzebną do osiągnięcia zakładanej dokładności

$$d = 2\,t_{a,n-1}\frac{s}{\sqrt{n}} \Rightarrow n \geq \left(2\,t_{a,n-1}\frac{s}{d}\right)^2$$

Choć n jest zmienną uwikłaną, to $t_{a,n-1}$ nie zmienia się szybko.

Podobnie można skonstruować przedział ufności dla populacyjnych charakterystyk rozproszenia: wariancji i odchylenia standardowego. Są one oparte o rozkład c^2 **– Pearsona** i mają postać:

$$P\left\{\frac{\mathrm{var}\,X}{c^2_{\frac{a}{2},\,n-1}} \le s^2 \le \frac{\mathrm{var}\,X}{c^2_{1-\frac{a}{2},\,n-1}}\right\} = 1 - a$$

$$P\left\{\sqrt{\frac{\mathrm{var}\,X}{c^2_{\frac{a}{2},\,n-1}}} \le s \le \sqrt{\frac{\mathrm{var}\,X}{c^2_{1-\frac{a}{2},\,n-1}}}\right\} = 1 - a$$

Określają one granice losowych (bo zależnych od próby losowej) przedziałów obejmujących nieznaną wartość wariancji i odchylenia standardowego w populacji.

Jeśli cecha **X ma rozkład dwupunktowy**, to jej charakterystyką jest ***p*** – wskaźnik struktury (**frakcja**).

Dysponujemy próbą: $x_1, x_2, \ldots, x_n$, gdzie $x_i = 1$ (‘sukces’) lub $x_i = 0$ (‘porażka’). $k = \sum_{i=1}^{n} X_i$ oznacza liczbę ‘sukcesów’.

Estymator punktowy frakcji *p*: $\hat{p} = \frac{k}{n}$

Przybliżony przedział ufności dla *p*:

$$\left(\hat{p} - u_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + u_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

gdzie poziom ufności **P = 1 –** a, a $\mathbf{u}_a$ jest kwantylem rzędu α rozkładu normalnego N(0, 1) **(**można go zastąpić wartością t(α, +∞)**)**.

Te przedziały ufności (dla p i dla σ^2) również mogą być jednostronne.

Przykład:

Badano stopień rozpowszechnienia telefonów komórkowych w środowisku studentów pewnej uczelni. Stwierdzono, że wśród 600 losowo przebadanych studentów telefon komórkowy posiadało 540 osób.

$$\hat{p} = \frac{k}{n} = \frac{540}{600} = 0{,}9$$

t(α, ∞), dla α = 0,05, ma wartość 1,96 (tablice statystyczne, rozkład t – Studenta).

Stąd zgodnie ze wzorem:

$$p \in \left(\hat{p} - u_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + u_{1-a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \quad P = 1 - a$$

$$p \in \left(0{,}9 - 1{,}96\sqrt{\frac{0{,}9(1-0{,}9)}{600}};\ 0{,}9 + 1{,}96\sqrt{\frac{0{,}9(1-0{,}9)}{600}}\right) P = 0{,}95$$

czyli

$$p \in \left(0{,}9 - 0{,}024\,;\,0{,}9 + 0{,}024\right) \;\; P = 0{,}95$$

$$p \in \left(0{,}876\,;\,0{,}924\right) \;\; P = 0{,}95$$

Z prawdopodobieństwem 0,95 mamy prawo oczekiwać, że prawdopodobieństwo posiadania przez pojedynczego studenta telefonu komórkowego będzie nie mniejsze niż 0,876, ale nie większe niż 0,924.

Z prawdopodobieństwem 0,95 **frakcja** studentów badanej uczelni, posiadających telefon komórkowy będzie nie mniejsza niż 87,6 % i nie większa niż 92,4 %.

Hipotezy statystyczne i ich weryfikacja, testy statystyczne

Drugim, obok estymacji (szacowania wartości parametrów lub rozkładu zmiennej losowej w populacji na podstawie rozkładu empirycznego dla próby), podstawowym rodzajem **wnioskowania statystycznego** jest **weryfikacja (testowanie) hipotez statystycznych**, czyli sprawdzanie określonych przypuszczeń (założeń) wysuniętych w stosunku do parametrów lub rozkładu populacji generalnej.

Hipotezy statystyczne to odpowiednio sformułowane przypuszczenia dotyczące rozkładu populacji. Mogą one mieć różną postać, w zależności od pierwotnych hipotez badawczych. Najczęściej stosuje się **hipotezy parametryczne**, precyzujące wartości parametrów populacyjnych (gł. średniej, wariancji czy frakcji).

Weryfikacja hipotezy statystycznej polega na stosowaniu specjalnego narzędzia, zwanego **testem statystycznym**. Jest to reguła postępowania, która każdej możliwej próbie losowej przyporządkowuje decyzję odrzucenia lub przyjęcia weryfikowanej hipotezy.

Istota każdego **testu** polega na tym, aby uchronić się przed popełnieniem **błędu I rodzaju** – polegającym na odrzuceniu hipotezy prawdziwej, jak i przed popełnieniem **błędu II rodzaju** – polegającym na nie odrzuceniu (czyli przyjęciu) hipotezy fałszywej.

Hipoteza $\mathbf{H_0}$	odrzucenie	przyjęcie
prawdziwa	α	$1 - \alpha$
fałszywa	$1 - \beta$	β

Jako poziom istotności a wybiera się najczęściej wartość: 0,05, choć można przyjąć dowolną liczbę z przedziału <0,1>.

W teorii weryfikacji (istotnościowych) hipotez statystycznych większe znaczenie przypisywane jest błędowi pierwszego rodzaju. Od testu statystycznego wymaga się, by błąd ten był ‘rzadko’ popełniany. Stąd zawsze narzuca się z góry pewne małe prawdopodobieństwo popełnienia błędu pierwszego rodzaju (**poziom istotności**, α) — ograniczające występowanie tego błędu.

Z testem statystycznym związane jest także pojęcie **mocy testu**.

Mocą testu nazywamy prawdopodobieństwo odrzucenia fałszywej hipotezy zerowej (lub prawdopodobieństwo nie odrzucenia hipotezy alternatywnej H_1, gdy w rzeczywistości jest ona prawdziwa). Prawdopodobieństwo to równe jest $1–\beta$.

Moc testu to inaczej prawdopodobieństwo nie popełnienia błędu drugiego rodzaju. Im większe jest to prawdopodobieństwo, tym lepszy jest dany test jako narzędzie do decydowania między hipotezą prawdziwą i fałszywą.

Test statystyczny może być słaby lub mocny:

- **test mocny** - w większości przypadków jesteśmy w stanie odrzucić fałszywą hipotezę zerową
- **test słaby** - gdy istnieje duża szansa na to, że nie odrzucimy hipotezy zerowej, pomimo jej nieprawdziwości.

Od testu powinniśmy wymagać, by był on jak ‘najmocniejszy’, tzn. by jak najłatwiej odrzucał hipotezę zerową, jeśli jest ona nieprawdziwa.

Test mocny:

1. rzadko myli się odrzucając H_0 (raczej nie odrzuca H_0 prawdziwej). Ustalane jest to za pomocą poziomu istotności
2. prezentuje mały błąd II rodzaju (β), czyli jeśli odrzuca H_0, to jest wysoka

szansa (równa mocy testu), że H_0 była fałszywa.

Wszystkie omawiane na wykładzie testy to testy w pewnym sensie najmocniejsze.

Hipoteza o średniej populacyjnej

Niech populacja generalna ma rozkład normalny $N(m,\sigma^2)$, przy czym oba parametry są nieznane. W oparciu o n – elementową próbę losową należy zweryfikować hipotezę zerową H_0: $m = m_0$, wobec hipotezy alternatywnej H_1: $m \neq m_0$

Dla weryfikacji tej hipotezy zerowej stosujemy test t – Studenta, który daje nam wartość statystyki temp obliczonej z próby:

$$t_{emp} = \frac{\bar{x} - m_0}{s}\sqrt{n} = \frac{\bar{x} - m_0}{s_{\bar{x}}}$$

Symbol $s_{\bar{x}}$ oznacza średni błąd średniej równy $S / n^{0,5}$.

Jeżeli zajdzie nierówność $|t_{emp}| \geq t_{\alpha,n-1}$, to hipotezę H_0 należy odrzucić na korzyść hipotezy alternatywnej H_1. Gdy zajdzie nierówność przeciwna, tzn. $|t_{emp}| < t_{\alpha,n-1}$, to nie mamy podstaw do odrzucenia hipotezy H_0, czyli przyjmujemy ją.

Jeśli cecha X ma rozkład $N(m;\sigma^2)$, to hipoteza zerowa: H_0: $m=m_0$ może być testowana różnie, w zależności od postaci hipotezy alternatywnej H_1.

Hipoteza alternatywna	Funkcja testowa	Obszar krytyczny
H_1: $m > m_0$	$t_{emp} = \frac{\overline{x}-m_0}{s_{\overline{x}}}$	$\left(t_{2a,n-1};+\infty\right)$
H_1: $m < m_0$	$t_{emp} = \frac{\overline{x}-m_0}{s_{\overline{x}}}$	$\left(-\infty;-t_{2a,n-1}\right)$
H_1: $m \neq m_0$	$t_{emp} = \frac{\overline{x}-m_0}{s_{\overline{x}}}$	$\left(-\infty;-t_{a,n-1}\right)\cup\left(t_{a,n-1};+\infty\right)$

Dwa pierwsze przypadki to hipotezy (i testy) jednostronne.

Przykład

Interesuje nas, czy średnia masa netto kubeczka jogurtu wynosi 200g. Pobieramy próbę prostą liczącą 20 kubeczków i określamy masę netto (np: x_1=191g , x_2=215g , x_3=201g ,... , x_{20}=189g). Dla hipotetycznej 20–elementowej próby przyjmijmy następujące wartości statystyk: średnia z próby 197,6, odchylenie standardowe w próbie wynosi 6.

$$t_{emp} = \frac{\bar{x} - m_0}{s}\sqrt{n} = \frac{197{,}6 - 200}{6}\sqrt{20} = -1{,}788$$

$$t_{0,05,19} = 2{,}093$$

Dla hipotezy alternatywnej H_1: $m \neq m_0$ wartość bezwzględna z t_{emp}=1,788 nie wpada do obszaru krytycznego ($|t_{emp}| > t_{a,v}$ 1,788 nie jest wieksze niż 2,093) więc: na poziomie istotności **5%** nie ma podstaw do **odrzucenia** hipotezy o tym, że **średnia** masa netto wynosi **200g**.

Gdybyśmy wykonywali test na poziomie istotności 10%, wtedy $t_{0,1,19}$=1,73, czyli warunek odrzucenia ($|t_{emp}|>t_{a,v}$) jest spełniony. Wtedy odpowiedź byłaby: na poziomie **istotności** 10% **odrzucamy** hipotezę zerową o tym, że **średnia** masa netto **wynosi 200g** na rzecz hipotezy alternatywnej, że **nie jest równa**.

Załóżmy, że hipotezę tę stawia biuro obrony konsumenta pod wpływem doniesień o zbyt małej wadze netto. Wtedy hipoteza zerowa ma postać H_0: $m = m_0$ a alternatywna H_1: $m < m_0$, tzn albo waga jest prawidłowa albo za mała. $t_{2*0,05,19}$=1,73 więc na poziomie istotności **5% odrzucamy** hipotezę zerową o tym, że **średnia** masa netto wynosi 200g na rzecz hipotezy alternatywnej, że **średnia** masa **jest mniejsza**.

Hipoteza o wartości wskaźnika struktury p

Zmienna losowa dwupunktowa, zero – jedynkowa.

$$H_0: p = p_0 \quad (H_1: p \neq p_0)$$

Statystyka empiryczna $z_{emp} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

gdzie $\hat{p} = \dfrac{k}{n}$

Jeśli $|z_{emp}| \geq u_{1-\alpha/2} = t_{\alpha,\infty}$, to hipotezę H_0 należy **odrzucić** na korzyść H_1. Natomiast, gdy $|z_{emp}| < t_{\alpha,\infty}$, to **nie mamy podstaw do odrzucenia** hipotezy H_0 (czyli merytorycznie uznajemy ją za prawdziwą).

Przykład:

Zweryfikujmy na **poziomie istotności** $\alpha = 0{,}05$ przypuszczenie, że wskaźnik wyposażenia studentów pewnej uczelni w laptopy jest równy 0.4, jeśli w losowej próbie 800 studentów fakt posiadania laptopa zadeklarowało 350 osób.

$$\hat{p} = \frac{k}{n} = \frac{350}{800} = 0{,}4375$$

$$H_0 : p = 0.4 \qquad z_{emp} = \frac{\left|\hat{p} - p_0\right|}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$z_{emp} = \frac{\left|0{,}4375 - 0{,}4\right|}{\sqrt{\frac{0{,}4(1-0{,}4)}{800}}} = 2{,}165$$

Ponieważ $t(0{,}05, +\infty) = 1{,}96$ to **odrzucamy** hipotezę zerową o tym, że **frakcja** w populacji wynosi **40%** (H_0: p = 40%) na korzyść hipotezy alternatywnej, że frakcja jest inna niż 40% (H_1: p ≠ 40%) przy przyjętym **poziomie istotności** równym 5%.

Można zaproponować jako alternatywną hipotezę jednostronną, np. H_1: $p > 0,4$, ale zmiana ta wpłynie na przebieg testowania - porównanie jednostronne z $u_{1-\alpha}$ czyli $t_{2\alpha,\infty}$.

Hipoteza alternatywna	Obszar krytyczny
H_1: $p > p_0$	$(u_{1-\alpha}, \infty)$
H_1: $p < p_0$	$(-\infty, u_{1-\alpha})$
H_1: $p \neq p_0$	$(-\infty, u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty)$

Hipotezy o wartości parametru a przedziały ufności dla tego parametru

Dla zadanego poziomu istotności α **hipoteza** o tym, że badany parametr populacyjny wynosi *x*, **jest odrzucana** wtedy i tylko wtedy, kiedy *x* **nie należy do przedziału ufności** skonstruowanego na poziomie ufności $1-\alpha$.